Development of an Authoring Tool that Promotes Accessibility

Phase I Performance Report

NIDRR SBIR Phase I

#H133S050060

Marucco, Stoddard, Ferenbach & Walsh, Inc.

Michael A. Scott, Project Director

April 7, 2006

Table of Contents

Development of an Authoring Tool that Promotes Accessibility Phase I Performance Report

Summary

MSF&W proposed the development of an "accessibility first" authoring tool that would enable non-technical authors to easily create accessible electronic documents. In Phase I, we examined the necessity and feasibility of this goal by testing two hypotheses:

(1) Existing authoring tools do not adequately support "typical" authors in producing accessible documents.

(2) An authoring tool designed to intentionally promote accessibility would significantly improve the likelihood that typical authors will produce accessible documents.

The project team developed a practical accessibility scoring metric, based on W3C and Section 508 standards, to measure and compare the accessibility of electronic documents. The team then developed a prototype "accessibility first" authoring tool with features designed to actively assist users in addressing accessibility. A series of user and expert tests were performed, comparing the accessibility of test documents created with existing authoring tools to those created using the "accessibility first" prototype.

In user testing, twelve non-technical users created a test document with three leading tools (Microsoft Word, Microsoft FrontPage, and Macromedia Contribute) and the prototype. The accessibility of each document was scored using the accessibility metric and compared across tools. Results were significant ($\alpha = .05$) and dramatic:

Microsoft Word, Microsoft FrontPage, Macromedia Contribute	**31.2%**
"Accessibility First" Prototype	**72.3%**

In expert testing, four accessibility experts created the test document using a wider range of authoring tools. Experts evaluated each tool in a "typical use" scenario, in which they mimicked the patterns of use observed in the user tests, and in a "best case" scenario, in which they did everything possible to create an accessible document. Results of "typical use" tests were similar to those found in user testing, showing an average of 38.7% accessibility for existing tools and 77.5% for the prototype. "Best case" tests demonstrated that accessibility was possible with many of these tools, with the best achieving as high as 95%. This observation led to an important additional conclusion:

It is not adequate to make accessibility ***possible***, it must be ***automatic***.

Based on the results of Phase I, the project team is convinced that it is both necessary and possible to develop a tool that would improve the accessibility of typical electronic documents. In Phase II, we will continue development of the prototype, with the goal of enabling a wide range of users – including those with disabilities – to regularly create documents that are over **95%** accessible.

Purpose

While accessibility guidelines have been in place for over ten years, many, if not most, organizations are still struggling to ensure that even simple electronic documents are adequately accessible to staff, customers, and constituents with disabilities. This problem stems from the fact that an increasing majority of electronic documents are created by non-technical authors who do not understand accessibility guidelines and the unfortunate reality that existing authoring tools do not encourage or assist users in addressing accessibility.

It is our belief that an authoring tool intentionally designed to promote accessibility will dramatically improve the accessibility of electronic documents created by typical authors. The purpose of Phase I was to test the two hypotheses that underlie this belief:

(1) Existing authoring tools do not adequately support "typical" authors in producing accessible documents.

(2) An authoring tool designed to intentionally promote accessibility would significantly improve the likelihood that typical authors will produce accessible documents.

The design of Phase I included four objectives:

(1) Identify a scoring metric that could be used to measure and compare the functional accessibility of electronic documents.

(2) Develop a prototype authoring tool that would automate and encourage users to address accessibility.

(3) Perform user tests to compare the accessibility of documents created by leading authoring tools versus the prototype.

(4) Perform expert evaluations to compare the accessibility of a wider range of authoring tools and the prototype in both "typical use" and "best case" scenarios.

The approach to and results of each objective are detailed in the following pages.

Project Team

Project Director - Michael Scott, M.S. Rehabilitation Engineering, a leading information technology accessibility expert in the State of Illinois and author of the Illinois Web Accessibility Standards, directed the project.

Researcher - Melissa Romanotto, M.A. Anthropology, a specialist in usability and user testing, designed and managed user and expert testing.

Programmer - Dustin Michaels, M.S. Computer Science, a specialist in accessible web development, developed the authoring tool prototype.

Consultant - Jon Gunderson, Ph.D., Director of Information Technology Accessibility at the University of Illinois at Urbana-Champaign and nationally-recognized information accessibility expert, assisted in the design of the prototype and testing.

Methods

Accessibility Scoring Metric

In establishing a scoring metric that could be used to measure and compare the functional accessibility of electronic documents, several specific requirements were identified:

(1) The metric should be based on the requirements and priorities of the W3C Web Content Accessibility Guidelines (WCAG) 1.0 and Section 508 standards for Web-based intranet and internet information and applications.

(2) The metric should apply specifically to electronic documents (as opposed to web sites), including elements such as text, images, tables, links, and formatting, but excluding multimedia, interactive elements (forms, scripts, applets), and web-site specific features (frames, navigation, etc).

(3) The metric should emphasize functional accessibility issues – i.e., those that affect how effectively a person with a disability can access essential information using screen magnification, screen reading, speech recognition, or other assistive tools.

(4) The metric should provide a quantifiable result that can be used for comparisons.

(5) The metric should be applicable to a test document that can be used in user testing.

Standards

The project team reviewed Section 508 and WCAG 1.0 standards and identified the specific standards and checkpoints that applied to electronic documents. Figure 1 and Figure 2 list the specific standards that were identified as applicable, including the type of page elements to which they apply:

Figure 1. Section 508 Standards Applicable to Electronic Documents

Standard	Requirement	Element
a	A text equivalent for every non-text element shall be provided (e.g., via "alt", "longdesc", or in element content).	Images
c	Web pages shall be designed so that all information conveyed with color is also available without color, for example from context or markup.	Color
d	Documents shall be organized so they are readable without requiring an associated style sheet.	Order
g	Row and column headers shall be identified for data tables.	Tables
h	Markup shall be used to associate data cells and header cells for data tables that have two or more logical levels of row or column headers.	Tables

Figure 2. WCAG 1.0 Checkpoints Applicable to Electronic Documents

Checkpoint	Priority	Requirement	Element
1.1	1	Provide a text equivalent for every non-text element.	Images
2.1	1	Ensure that all information conveyed with color is also available without color.	Color
3.1	2	When an appropriate markup language exists, use markup rather than images to convey information.	Code
3.2	2	Create documents that validate to published formal grammars.	Code
3.3	2	Use style sheets to control layout and presentation.	Code
3.4	2	Use relative rather than absolute units in markup language attribute values and style sheet property values.	Size
3.5	2	Use header elements to convey document structure and use them according to specification.	Headings
3.6	2	Mark up lists and list items properly.	Lists
3.7	2	Mark up quotations. Do not use quotation markup for formatting effects such as indentation.	Code
5.1	1	For data tables, identify row and column headers.	Tables
5.2	1	For data tables that have two or more logical levels of row or column headers, use markup to associate data cells and header cells.	Tables
5.3	2	Do not use tables for layout unless the table makes sense when linearized.	Order
5.4	2	If a table is used for layout, do not use any structural markup for the purpose of visual formatting.	Code
6.1	1	Organize documents so they may be read without style sheets.	Order
11.1	2	Use W3C technologies when they are available and appropriate for a task and use the latest versions when supported.	Code
11.2	2	Avoid deprecated features of W3C technologies.	Code
13.1	2	Clearly identify the target of each link.	Links

Elements & Weighting

To provide appropriate emphasis for key elements, a simple weighting scheme was established:

Figure 3. Weighting Scheme

Elements	Weight
Elements relating to WCAG Priority 2 Checkpoints	1
Elements relating to WCAG Priority 1 Checkpoints and Section 508 Standards	2

This weighting scheme is similar to approaches taken by other researchers, such as the Web Accessibility Barriers (WAB) scoring method proposed by Xiaoming Zeng, M.D., Ph.D. (Zeng, X. 2004. Evaluation and Enhancement of Web Content Accessibility for Persons with Disabilities. Dissertation. University of Pittsburgh) which assigned points to WCAG Checkpoints based the inverse of the WCAG Priority level.

This scheme also fit well with the project team's expectations based on practical experience gained in training and supporting computer users with disabilities: It placed appropriate emphasis on images, tables, color and order while also including key elements that are practically significant but often overlooked, such as headings, lists, links, size, and code.

Based on the weighting scheme, elements were weighted as follows:

Figure 4. Elements & Weighting

Element	Criteria	Standards	Weight
Images	Simple images have appropriate alt text; complex images have full descriptions	WCAG 1.1; 508 a	2
Tables	Simple tables have row and column headers; complex tables have cell/header associations	WCAG 5.1, 5.2; 508 g, h	2
Headings	Headings form table of contents or outline	WCAG 3.5	1
Lists	List type and nesting can be identified	WCAG 3.6	1
Links	Link text is understandable out of context	WCAG 13.1	1
Size	Text can easily be resized	WCAG 3.4	1
Color	No information provided by color alone	WCAG 2.1; 508 c	2
Order	Reading order is correct when tables are linearized and style sheets are ignored	WCAG 5.3, 6.1; 508 d	2
Code	Valid HTML & CSS code	WCAG 3.1, 3.2, 3.3, 3.7, 11.1, 11.2	1

Test Document

Test Elements

A test document (shown in Figure 6 and Figure 7) was developed, incorporating as many of the identified elements as possible in realistic sample document. Multiple test elements were included for each category in order to provide a complete representation of the all the likely scenarios in which each element could be used. For example, three separate images were included, representing:

(1) A simple image, requiring basic alternate text

(2) A decorative image, requiring empty alternate text (alt="")

(3) A complex image, requiring alternate text and a full description

Points

To facilitate scoring, points were assigned to each element category according to its weight. Specifically, element categories with weight 1 were allotted 10 points and elements with weight 2 were allotted 20 points. These points were then distributed between the test elements as shown below:

Figure 5. Test Document Elements

Element	Weight	Points	Test Document Elements
Images	2	20	3 images: 1 simple (5 points), 1 decorative (5 points), 1 complex (10 points)
Tables	2	20	2 tables: 1 simple (10 points), 1 complex (10 points)
Headings	1	10	5 headings (2 points each)
Lists	1	10	1 bulleted list, 1 numbered list, 3 sub-lists (2 points each)
Size	1	10	All text in document (20 elements; .5 point each)
Code	1	10	HTML & CSS code (5 points each)

Exclusions

For practical reasons, links, color, and order were excluded from the test document.

Because of the nature of the user test – in which users would be creating electronic documents based on a printed original – it was not possible to prompt users to create links or use color without being completely prescriptive; based on what was shown in the printed original, users would either get these elements all right or all wrong.

Order was initially included in the test document through the use of a two-column newsletter-style layout. However, in pilot testing, users had so much difficulty attempting to use the column layout features of the authoring tools that they were unable to complete the tests within the four-hour time limit.

Figure 6. Test Document, Page 1

U.S. Census Data for Midwestern States

The official U.S. Census is described in Article I, Section 2 of the Constitution of the United States. It calls for an actual enumeration of the people every ten years, to be used for apportionment of seats in the House of Representatives among the states. The first official Census was conducted in 1790. Since that time, the decennial Census has been conducted every ten years, generally on April 1 in years ending in a zero.

Besides providing the basis for congressional redistricting, Census data are used in many other ways. Since 1975, the Census Bureau has had responsibility to produce small-area population data needed to redraw state legislative and congressional districts. Other important uses of Census data include:

- the distribution of funds for government programs such as Medicaid
- planning the right locations for schools, roads, and other public facilities
- helping real estate agents and potential residents learn about a neighborhood
- identifying trends over time that can help predict future needs.

Besides the decennial census, the Census Bureau conducts nearly one hundred other surveys and censuses every year. The information collected for each census and survey is summarized by geographic area and then published in a variety of formats, including printed reports, CD-ROM, DVD, and on the Internet.

The American Community Survey (ACS) is a new nationwide survey designed to provide communities a fresh look at how they are changing. It is intended to eliminate the need for the long form in the 2010 Census. The ACS collects information from U.S. households similar to what was collected on the Census 2000 long form, such as income, commute time to work, home value, veteran status, and other important data.

The ACS began in 1996 and has expanded each subsequent year. Data from the 2004 ACS are available for over 800 geographies, including 244 counties, 203 congressional districts, most metropolitan areas of 250,000 population or more, all 50 states, and the District of Columbia.

Data collected includes:

A. Total housing units
    1. Occupied housing units
        a. Owner-occupied housing units
        b. Renter-occupied housing units
            i. Gross rent
            ii. Gross rent as percentage of household income
    2. Vacant housing units
B. Average household size
C. Average family size

Population Characteristics

Population by Sex

Between 1990 and 2000, the male population grew slightly faster than the female population. The excess of the female to male population dropped to 5.3 million in 2000. This resulted in the male-female ratio increasing from 95.1 in 1990 to 96.3 in 2000.

Figure 7. Test Document, Page 2

	Males (in millions)	Females (in millions)	Total (in millions)
Illinois	6.1	6.3	12.4
Indiana	3	3.1	6.1
Iowa	1.4	1.5	2.9
Missouri	2.7	2.9	5.6
Wisconsin	2.7	2.7	5.4

Population by Age & Sex

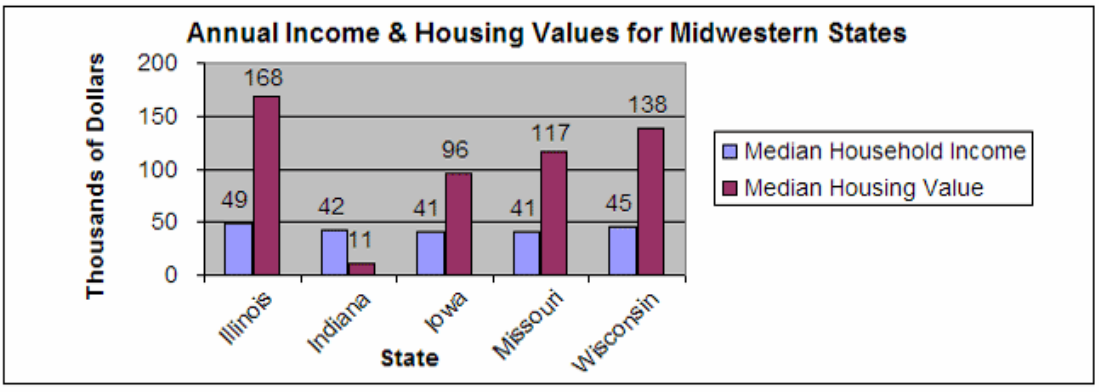
In general, the ratio of the male population to the female population declined with age. That is, the female population exceeded the male population at older ages, but the reverse was true at younger ages. In 2000, there were 20.6 million women aged 65 and over compared with only 14.4 million men. In contrast, there were 37.1 million males under 18 while there were 35.2 million females. The balance shifted toward more women at age 36.

	Males (in millions)	Females (in millions)	Total (in millions)
Indiana			
18 Years and Older	2.2	2.3	4.5
65 Years and Older	0.3	0.4	0.7
Missouri			
18 Years and Older	2	2.2	4.2
65 Years and Older	0.3	0.4	0.7
Wisconsin			
18 Years and Older	2	2	4
65 Years and Older	0.4	0.3	0.7

Annual Income & Housing Values

Median income remained unchanged for all types of family and non-family households (such as married-couple households and single individuals) between 2002 and 2003. The real median earnings of men who worked full-time, year-round remained at $40,668. The real median earnings of the comparable group of women declined to $30,724. Reflecting the fall in earnings of women, the female-to-male earnings ratio declined from 0.77 to 0.76 between 2002 and 2003.

Two thirds of the households in the U.S. lived in a home of their own in 2000. Almost 80 percent of owner-occupied housing units were single-family homes. The median value was $119,600.



Scoring Guide

A scoring guide (Figure 8 & Figure 9) was developed to specify how the elements in the test document would be quantitatively evaluated. Each test element was assigned a base point value based on the weighting scheme. Likely errors for each element were identified, and penalty values assigned according to severity of the error from an accessibility perspective (i.e., the more information that would be "lost" to an assistive technology user, the more severe the error and the penalty).

For example, the omission of alternate text for a meaningful image would effectively make all of that image's information unavailable to an individual using a screen reader, so the penalty for that error deducts all points for that element. In the case of alternate text that is present but incomplete, the penalty deducts a portion of that element's points.

For elements containing important structural information (e.g., a table or a heading) penalties were based on the presence or absence of appropriate indication of the appropriate structure. For example, a heading that was formatted as bold instead of as a heading would be missing its structural information and would be penalized accordingly.

Scoring

Scoring was performed by reviewing completed test documents against the scoring guide. Each test element was checked to determine whether any of the errors identified in the guide were present. If an error was found, the corresponding penalty was deducted from the point value of the element, and the resulting score was recorded. In the case of multiple errors, only the largest penalty was used. When review of the entire document was completed, the awarded points were totaled and divided by the total possible points for the document, providing a final accessibility score ranging from 0 to 100%.

A Microsoft Access database application was developed to automate scoring calculations. An Access form was used to collect element scores for each document, and the application performed all the necessary calculations to analyze the data.

Document Formats

It is important to note that while the scoring metric focuses on functional accessibility issues, the details of the scoring guide are presented in HTML-oriented terms. Most of the tools tested used HTML as their native document format; for the tools that used other native formats (Microsoft Word, WordPerfect, and OpenOffice), documents were saved as HTML using the best available method (i.e., "Filtered" HTML in Word, "Plain HTML" in WordPerfect) for scoring.

Figure 8. Scoring Guide

Test Element	Case	Points	Penalty
Image 1 (simple)	Correct (alt="United States Census 2000")	5	
	Excessive alt text (too descriptive, “image” words)		-1
	Inadequate alt text (missing words)		-3
	Misleading alt text (other text, alt="", no alt attribute)		-5
Image 2 (decorative)	Correct (alt="")	5	
	Excessive alt text (descriptive)		-1
	Inadequate alt text (no alt attribute)		-3
	Misleading alt text (other text)		-5
Image 3 (complex)	Correct (alt=”Annual Income & Housing Values” & complete narrative or data table for full description)	10	
	Excessive alt text (too descriptive, “image” words)		-1
	Inadequate alt text (missing words)		-3
	Misleading alt text (other text, alt="", no alt attribute)		-5
	Excessive full description		-1
	Inadequate full description		-3
	Missing or misleading full description		-5
Table 1 (simple)	Correct (row & column headers)	10	
	Column headers only (no row headers)		-2.5 *
	Row headers only (no column headers)		-2.5 *
	No row or column headers		-5 *
	No table		-10
Table 2 (complex)	Correct (column & row headers, headers & ids)	10	
	Missing headers attribute on row headers		-1 *
	Column & row headers only (no headers or ids)		-4 *
	Column headers only (no row headers, headers, ids)		-5.5 *
	Row headers only (no column headers, headers, ids)		-5.5 *
	No row or column headers, headers or ids		-7 *
	No table		-10

Figure 8. Scoring Guide, continued

Element	Case	Points	Penalty
Headings	Correct (5 headings: h1, h2, h3, h3, h2)	10	
	Heading out of order		-1 each
	Heading missing or misused		-2 each
Lists	Correct (1 bulleted list, 1 numbered list, 3 sub-lists)	10	
	Improper list type or nesting		-1 each
	List missing or misused		-2 each
Font Size	Correct (all text with relative font size)	10	
	Elements with absolute font size (pt, px, in, etc)		-.5 each
Code	Valid HTML 4.01 or XHTML 1.0	10	
	Missing or extraneous attributes present		-1
	Non-standard elements present		-3
	Not well-formed HTML		-5
	HTML elements used for styling (b, big, center, font, i, s, small, strike, tt, u, etc)		-1 each type

* Table scores are determined by evaluating the number of correctly identified cell/header relationships compared to the number of necessary cell/header relationships for each sample table using the following guidelines:

Figure 9. Table Scoring Detail

Element	Case	Points	Penalty
Table cell/header relationship	Cell associated with header by position or headers	2 each	
	Not explicit (header not indicated with markup)		-1 each
	Incorrect or missing relationship		-2 each

Element	Possible “Relationship” Points
Table 1	15 data cells x 2 relationships x 2 points = 60 possible relationship points
Table 2	18 data cells x 3 relationships x 2 points + 6 row headers x 1 relationship x 2 points = 120 possible relationship points

Table Score	Calculation
Table Score	Actual relationship points / possible relationship points x 10

Authoring Tool Prototype

Functional Requirements

The Phase I authoring tool prototype was designed to address the priorities identified in the accessibility scoring metric. Wherever possible, accessibility is automated; where necessary, prompts and instructions are provided to guide the user in addressing subjective accessibility issues.

The specific functional requirements of the authoring tool prototype included:

Figure 10. Authoring Tool Functional Requirements

Element	Authoring Tool Features
Images	Require alternate text; prompt and instruct user on appropriate use. Provide "image check" tool to review appropriateness. Prompt user to identify complex images (charts, graphs, etc). Prompt and instruct user on appropriate use of full description.
Tables	Automatically use table header cells in first column and first row. Automatically add id and headers attributes when table cells are merged. Provide "table check" to visualize and confirm header/cell associations.
Headings	Offer easy options for headings; do not allow style-based formatting. Provide "heading check" to visualize and confirm order of headings.
Lists	Offer easy options for structural lists. Provide "list check" to confirm proper use of structural lists.
Size	Automatically use relative font sizes.
Code	Automatically produce valid XHTML 1.0 and CSS code.

Technical Design

The prototype was built as a web application using the MSHTML Editing Platform, which is a customizable WYSIWYG editing control that generates and renders HTML code. Because it does not natively produce valid HTML code and many of its built-in features violate accessibility requirements, the control was extensively customized to meet the functional requirements. Custom screens and functionality were developed using XHTML 1.0, CSS 1 & 2, and JScript 5.5. A simple ASP.NET 1.1 and SQL Server 2000 content management tool was used to facilitate saving and storing test documents.

As a web application, the prototype was designed according to WCAG Priority 1 & 2 guidelines and Section 508 standards; however, it is not yet fully usable with screen reading or speech recognition tools. Enhancing the accessibility of the prototype is a primary goal of Phase II.

Prototype

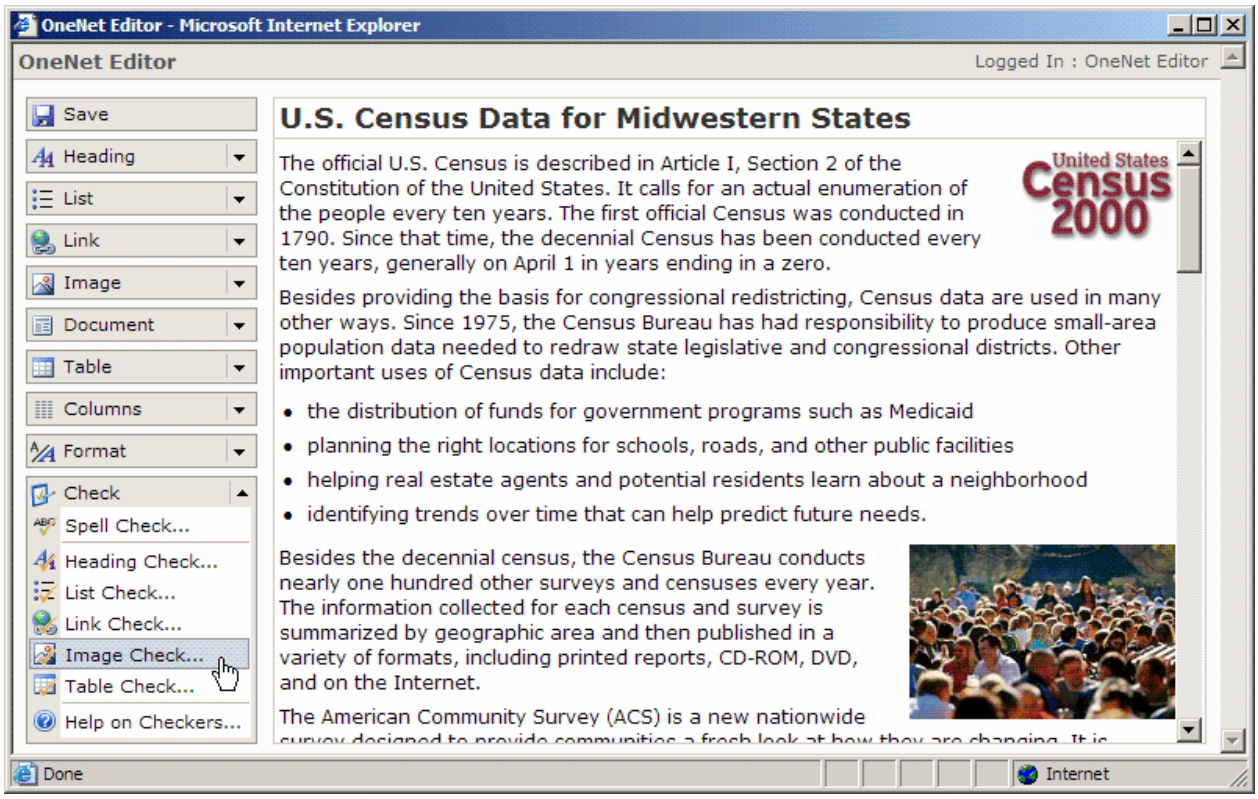
To facilitate usability, the prototype was intentionally designed with a simple, minimalist user interface. Accessibility features and instructions were incorporated throughout, making accessibility an integral part of the authoring experience rather than an afterthought.

Editor Screen

The editor screen displays the primary functions of the prototype in a series of collapsible menus (left side). All menus are labeled with a simple icon and a textual name to facilitate easy recognition. The order of the menus is designed to emphasize often overlooked accessibility features, such as headings and lists, and de-emphasize often misused features, such as bold and indent (which are found within the Formatting menu). The most problematic features, such as font-based formatting options, are intentionally omitted.

The editing area (right side) is designed to function very much like a typical word processing program. Users are able to enter and edit text, insert images, create and modify tables, add formatting, and interact with their documents in a WYSIWYG (what-you-see-is-what-you-get) view.

Figure 11. Authoring Tool Prototype Editor Screen



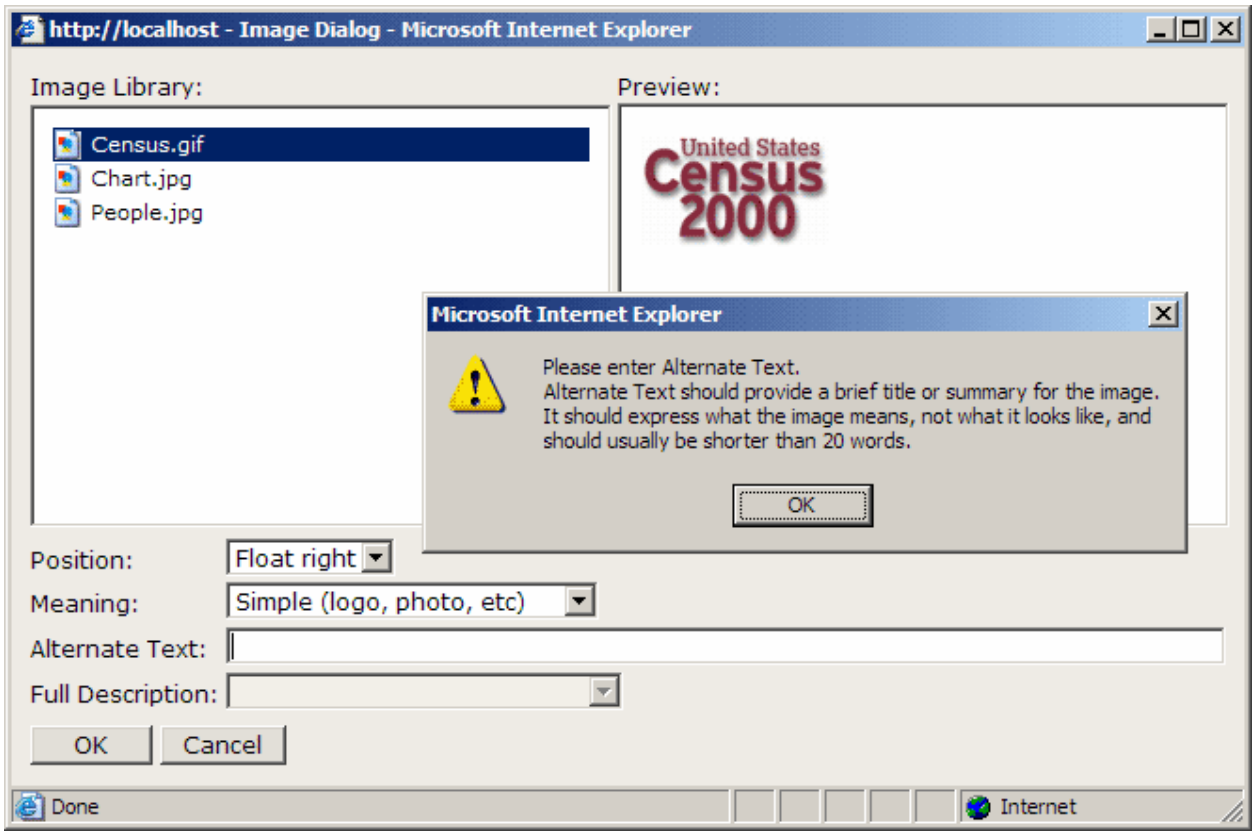
Dialog Boxes

Dialog boxes, such as the Image Dialog, are provided for editing functions that require users to make choices or provide additional information. The Image Dialog is used to add an image to a document. When opened, the Image Dialog displays a library of images, allowing the user to preview each image and make the desired selection. Entry fields for key accessibility-related information, such as alternate text and full description, are included prominently on the screen.

Instructions and guidance are integrated into the dialog. A "meaning" field prompts users to identify the meaning of the image: "none (decorative only)", "simple (logo, photo, etc)", or "complex (chart, graph, etc)". If "none" is selected, the alternate text and full description fields are disabled (and alt="" is automatically used); if "simple" is selected, the alternate text field is activated and required; and, if "complex" is selected, both fields are activated and required.

The image dialog also provides intelligent prompting. If a user forgets to enter alternate text, the dialog presents a warning message and provides instruction in layperson's terms regarding the proper use and form of alternate text. Similarly, the Image Dialog checks for common mistakes such as alternate text that is too short (fewer than 2 words), too long (more than 20 words), or contains redundant terms ("image", "graphic", etc) and instructs the user on how to make corrections.

Figure 12. Authoring Tool Prototype Image Dialog



Automation

Wherever possible, the prototype automates accessibility without requiring any action from the user. For example, when a user creates a table, the cells in the first row and first column are automatically identified as header cells. If a user merges cells in a row, the appropriate id and headers attributes are calculated and added.

Similarly, code generated by the MSHTML Editing control is automatically cleaned and re-formatted so that it complies with XHTML 1.0 and CSS standards. (HTML Tidy is used in the final stage of the re-formatting process.) All code-related functions are performed exclusively behind-the-scenes; the user is never asked to view or interact with code.

Checks

The prototype also includes a series of accessibility "checks" designed to help users visualize and confirm subjective accessibility issues. These checks are available at any time, and the system reminds the user to run them whenever the document is saved.

Heading Check

Heading Check (Figure 13) displays all the headings in the document in an ad-hoc "table of contents" view. The user is prompted to review the list to confirm that it represents the intended organization of the document. A warning is given if headings are used out of sequence or if it appears that bold formatting has been used where headings were intended. The user can make any necessary modifications to the text or level of the headings directly in the Heading Check window.

List Check

List Check (Figure 14) scans the document for "fake lists" – lists that are created using symbols or numbers rather than the List menu options. If fake lists are found, they are displayed in the List Check window and the user is prompted to convert them to actual lists.

Image Check

Image Check (Figure 15) displays a list of all the images in the document, showing the image, its meaning type, its alternate text, and a link to its full description (if one exists). The user is prompted to scan the list and confirm that the alternate text and full descriptions appropriately convey the meaning of the image.

Table Check

Table Check (Figure 16) displays each table in the document in a linearized view, indicating the table's caption, summary, and the headings associated with each cell. The user is prompted to review this list to confirm that tables are presented appropriately.

Figure 13. Heading Check Dialog

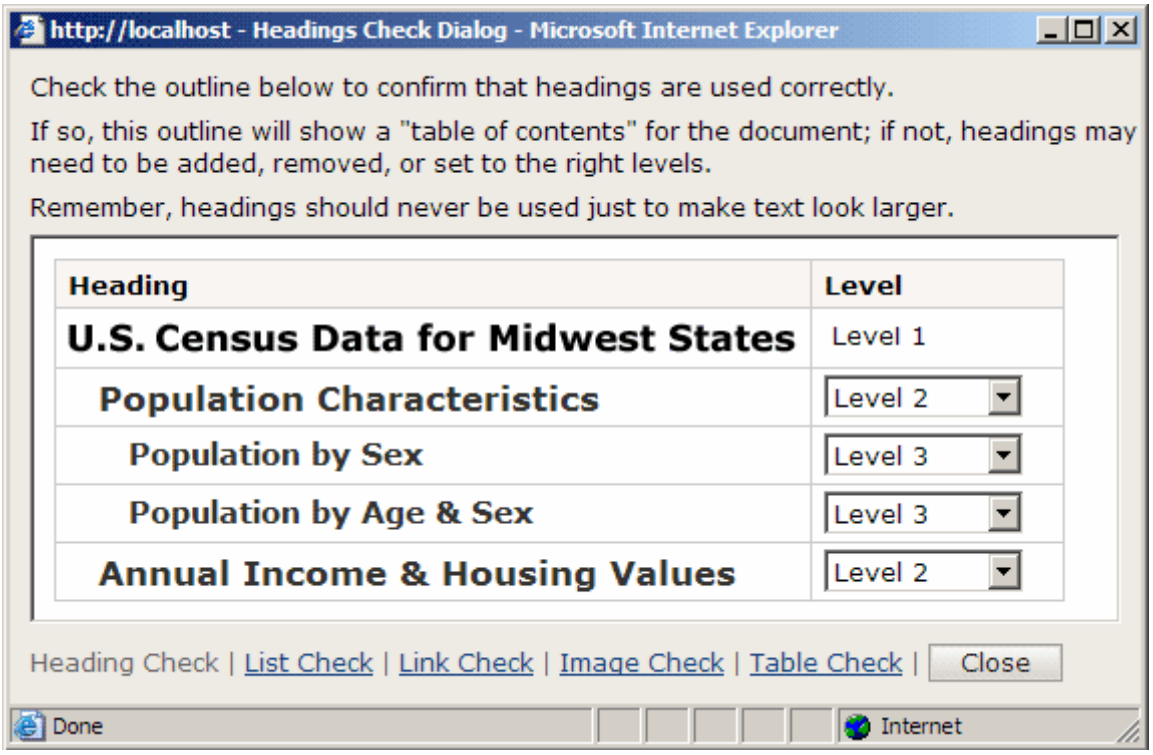


Figure 14. List Check Dialog

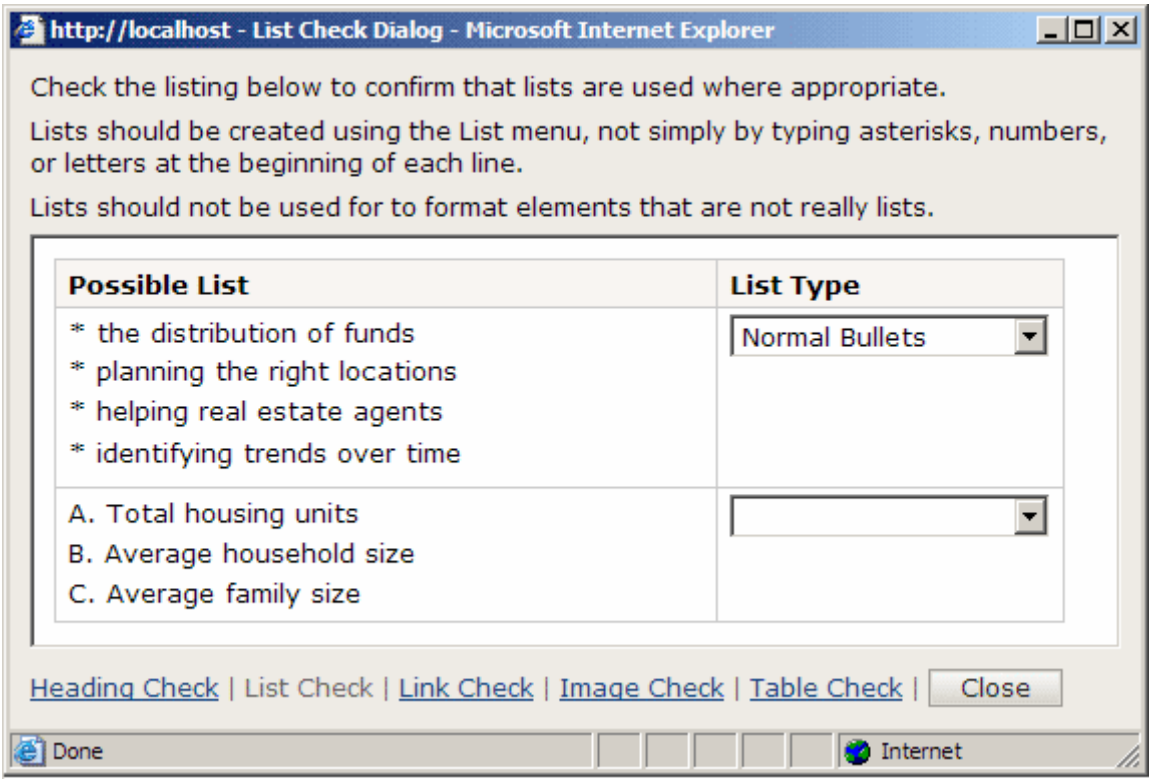


Figure 15. Image Check Dialog

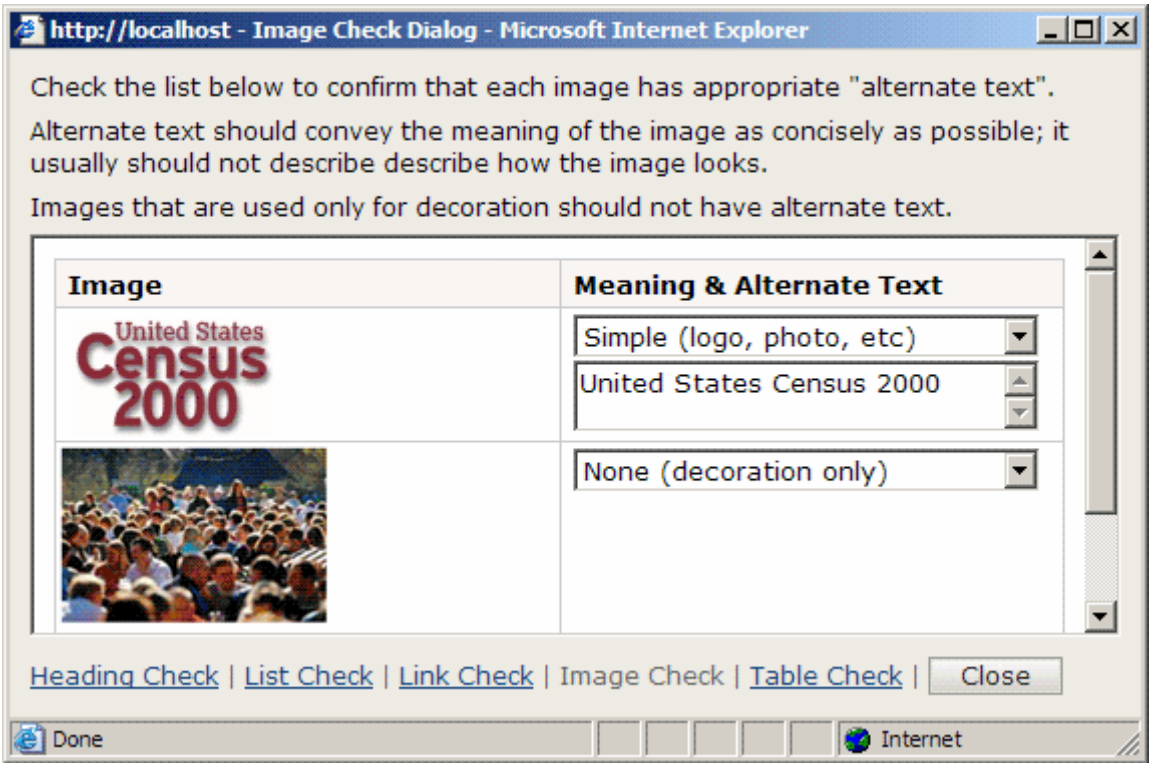
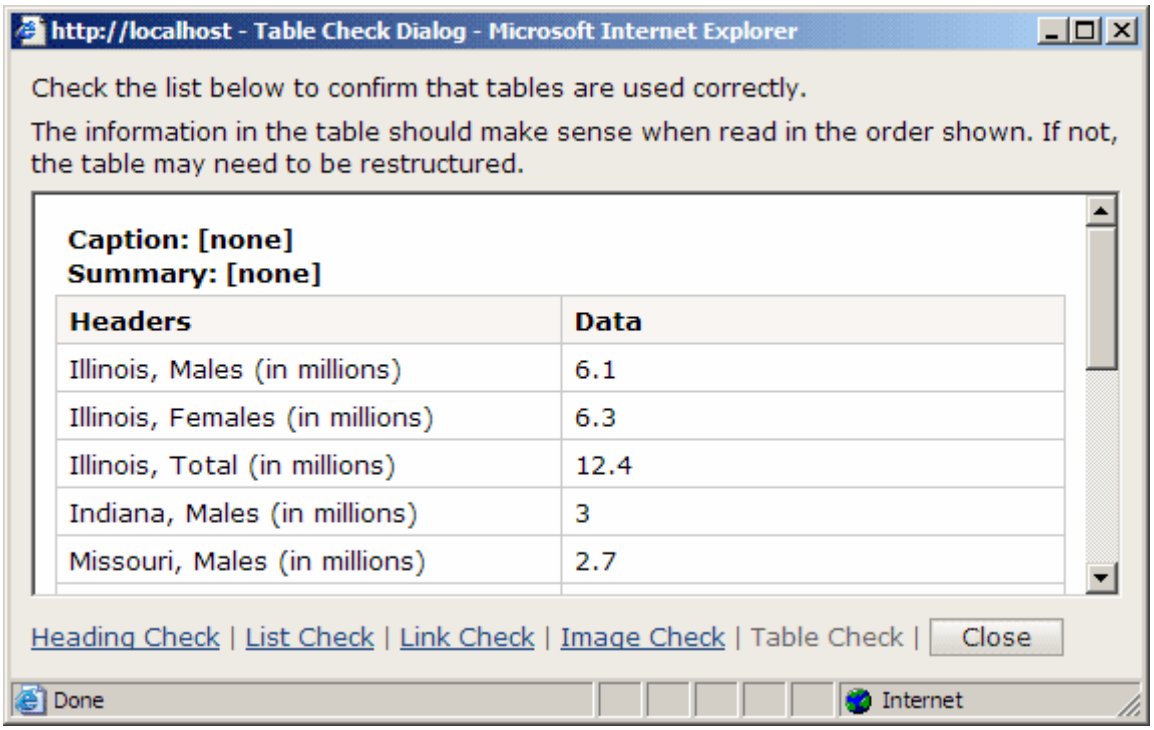


Figure 16. Table Check Dialog



User Tests

The user tests were designed to objectively evaluate and compare how well authoring tools enabled and encouraged typical "non-technical" users to create accessible documents.

Twelve users were asked to re-create a printed copy of the test document using each of four authoring tools – Microsoft Word, Microsoft FrontPage, Macromedia Contribute, and the "accessibility first" prototype. Each resulting electronic document was collected and evaluated using the accessibility metric scoring guide. Scores were totaled, converted to percentages (by dividing by the total possible points in the document), averaged, and compared.

Participants

User testing required twelve typical "non-technical" users. These users were defined as:

(1) Being comfortable using word processing software

(2) Not being information technology professionals or "expert users"

(3) Having little or no familiarity with information accessibility

Participants were recruited through flyers posted at a local community college and directed to contact the Project Researcher by telephone. During those calls, the Researcher administered a screening questionnaire to confirm that the user appeared to match the "non-technical" profile. Care was taken to avoid revealing that the test involved accessibility; Users were instead told that the test was intended to compare the ease-of-use of four authoring tools.

Before each test session, users completed a written background questionnaire in which they answered the following questions for each of the three existing authoring tools (Word, FrontPage, and Contribute):

(1) Approximately how many years have you been using [tool name]?

(2) Approximately how many hours per week do you work with [tool name]?

(3) Please rate your level of proficiency with [tool name], using a scale from 1 (beginner) to 5 (expert).

After finishing the test session, users completed a written post-test questionnaire in which they were asked:

(4) Please rate [tool name] on ease of use, using a scale from 1 (very difficult) to 5 (very easy).

(5) Before this testing session, were you familiar with web accessibility for people with disabilities? Use a scale from 1 (No) to 5 (I am an expert).

Demographic profiles of each of the twelve user test participants, along with their answers to key pre- and post-test questionnaires, are shown in Figure 17.

Figure 17. User Test Participants

User	Profession	Age	Sex	Experience w/ Word (years)	Proficiency w/ Word (1-5)	Familiarity w/ Accessibility (1-5)
1	Student	18-25	F	7.5	3	1
2	Student	18-25	F	8	3	1
3	Student	50-59	F	2	2	1
4	Student	18-25	M	6.5	4	1
5	Student	30-39	F	14	3	1
6	Clerical	40-49	F	10	3.5	2
7	Professional	40-49	M	14	4	1
8	Clerical	40-49	F	10	3.5	1
9	Student	18-25	M	4.5	3	2
10	Professional	30-39	F	10	2	1
11	Professional	30-39	M	10	4	1
12	Professional	40-49	F	15	3	2
Average		30-39		9.3	3.2	1.3

Tools

Three “leading” authoring tools were selected based on their popularity and their intended target audience. Microsoft Word was selected as the most widely-used authoring tool. Microsoft FrontPage was selected as a web design tool popular among beginning and intermediate web authors that also includes specific accessibility-related features. Macromedia Contribute was selected as the “non-technical” counterpart to the industry-leading Dreamweaver web development tool.

Figure 18. User Test Tools

Tool	Type	Audience
Microsoft Word 2003	Word processor	Non-technical
Microsoft FrontPage 2003	Web design tool	Technical
Macromedia Contribute 3.11	Web page editor	Non-technical
Prototype	Electronic document editor	Non-technical

Tests

Three user test sessions, each with four users, were conducted between March 2 and March 4, 2006. Each test session lasted four hours and was held in the Assistive Technology Testing & Training Laboratory in Springfield, Illinois. The Project Researcher served as test monitor and the Project Director observed the user tests.

In each test session, users were given a printed copy of the test document and instructed to re-create it four times, once in each of the four authoring tools in the order specified on an instruction sheet. To reduce the burden of typing, users were provided an initial document for each tool containing the unformatted text of the test document, except for the content of the tables. Users were instructed to format the provided text or re-create it as they saw fit. They were also instructed to apply a "normal" level of effort to the task and to avoid going to any unusual lengths to accomplish any part of the task that they would not normally attempt. No assistance was provided to the users during the tests.

Test Order

To minimize the potential for the order in which the tools were tested to affect the results, the testing order was carefully controlled. Each user was given a specific sequence to follow so that each tool was used in each position three times, and each tool immediately preceded/followed each other tool three times. The test order for each user is detailed below:

Figure 19. User Test Order

User	1st	2nd	3rd	4th
1	Word	FrontPage	Contribute	Prototype
2	Word	Contribute	Prototype	FrontPage
3	Word	Prototype	FrontPage	Contribute
4	FrontPage	Word	Prototype	Contribute
5	FrontPage	Contribute	Word	Prototype
6	FrontPage	Prototype	Contribute	Word
7	Contribute	Word	FrontPage	Prototype
8	Contribute	FrontPage	Prototype	Word
9	Contribute	Prototype	Word	FrontPage
10	Prototype	Word	Contribute	FrontPage
11	Prototype	FrontPage	Word	Contribute
12	Prototype	Contribute	FrontPage	Word

Expert Tests

Expert tests were designed to compliment the user tests by examining a wider range of authoring tools and by providing a comparison between what was likely in "typical use" and what would be possible in the "best case" scenario for each tool.

In informal testing sessions, the project team accessibility experts re-created the test document using the four authoring tools used in user testing along with six additional tools. For each tool, the expert testers created the test document twice, once in a "typical use" scenario, in which they mimicked the patterns of use observed in the user tests, and again in a "best case" scenario, in which they did everything possible to create an accessible document, even if it required activating special features of the authoring tool or applying special knowledge of accessibility that typical users would not possess. Each resulting document was evaluated using the accessibility metric scoring guide. Scores were totaled, converted to percentages, averaged, and compared, in the same method as the user tests.

Tools

In addition to Word, FrontPage, Contribute, and the prototype, expert tests evaluated six additional tools: Corel WordPerfect and OpenOffice.org Writer were evaluated as alternatives to Microsoft Word. Macromedia Dreamweaver was evaluated as an industry-leading web development tool with extensive accessibility-oriented features. eWebEditPro, r.a.d.editor, and XStandard were evaluated as small-market, web-based web editing tools similar in construction to the prototype, each of which had made claims of accessibility or Section 508 compliance.

Figure 20. Expert Test Tools

Tool	Type	Audience
Microsoft Word 2003	Word processor	Non-technical
Corel WordPerfect X3	Word processor	Non-technical
OpenOffice.org Writer 2.0.1	Word processor	Non-technical
Microsoft FrontPage 2003	Web design tool	Technical
Macromedia Dreamweaver 8.0.1	Web design tool	Technical
Macromedia Contribute 3.11	Web page editor	Non-technical
Ektron eWebEditPro 4.2.0	Web page editor	Non-technical
Telerik r.a.d.editor 5.64	Web page editor	Non-technical
XStandard Pro 1.6.2	Web page editor	Technical
Prototype	Electronic document editor	Non-technical

Scenarios

Typical Use

To guide expert testing in the "typical use" scenario, qualitative observations of user behaviors during the user tests were compiled into the following typical use guidelines. Expert testers followed these guidelines strictly while performing the "typical use" tests:

Figure 21. Typical Use Guidelines

Element	Guidelines
Simple Images	Do not enter alt text unless it is required.
Complex Images	Do not enter a full description unless it is required. Do not enter a full description correctly unless given clear instructions.
Simple Tables	Use formatting (bold, center, background color) for header cells. Do not select table headers unless the option is prominent and easy to use (e.g., on the table dialog and can be set for entire table at once).
Complex Tables	Do not select or edit id or headers attributes.
Headings	Use formatting (font size, bold) if formatting options are available. Do not use headings or heading styles unless prompted to do so.
Lists	Use toolbar list options for bulleted lists if available. Type numbers/letters for numbered lists. Use tab to indent sub-lists.
Size	Select all text and set font size using most readily available option. Use points or pixels unless discouraged.

Best Case

In the "best case" scenario, experts were instructed to do everything possible, short of editing HTML code directly, to achieve perfect accessibility scores with each tool. Knowing the intended outcome for each of the elements in the test document, experts were directed to seek any and all means of achieving the desired results. Experts were also instructed to utilize any accessibility validation or visualization utilities that were available in the tools.

Results

User Test Results

Use test results strongly supported the two hypotheses proposed in this study:

(1) Existing authoring tools do not adequately support "typical" authors in producing accessible documents.

(2) An authoring tool designed to intentionally promote accessibility would significantly improve the likelihood that typical authors will produce accessible documents.

As summarized in Figure 22, the existing tools tested averaged only 31.2% on the accessibility scoring metric. Given that the metric is a general measure of the amount of information that is presented in an accessible manner, it is apparent that these tools are not adequately supporting accessibility. Conversely, the "accessibility first" prototype shows an average score of 72.3%, clearly a significant improvement over the existing tools.

Figure 23 and Figure 24 depict the distribution of user scores for each tool, showing that results were consistent across users. The box plot in Figure 24 shows that the average score for the prototype was well outside of the +/- 1 standard deviation ranges of for the other tools. ANOVA calculations (Table A-5 & Table A-6 in the Appendix) confirm the difference between the scores for the prototype and the other tools were statistically significant at the 95% confidence level.

Encouragingly, the prototype achieved a maximum score of 98.8%, suggesting that it may be possible to enable typical users to achieve almost perfect accessibility. Also worthy of note is the fact that the lowest score for the prototype (53.8%) was higher than the highest score of any of the other tools (44.4% for FrontPage and Contribute).

Discouragingly, the most widely used tool, Microsoft Word, was by far the worst performer, averaging only 22.5% and peaking at 25%.

Figure 22. Average User Test Accessibility Scores (%)

Tool	Average Score	Minimum Score	Maximum Score	Standard Deviation	Number of Testers
Word	22.5	17.5	25	1.9	12
FrontPage	35.9	28.1	44.4	3.9	12
Contribute	35.3	16.9	44.4	8.7	12
Word, FrontPage, Contribute	31.2	16.9	44.4	8.3	36
Prototype	72.3	53.8	98.8	12.0	12

Figure 23. User Test Accessibility Scores (%)

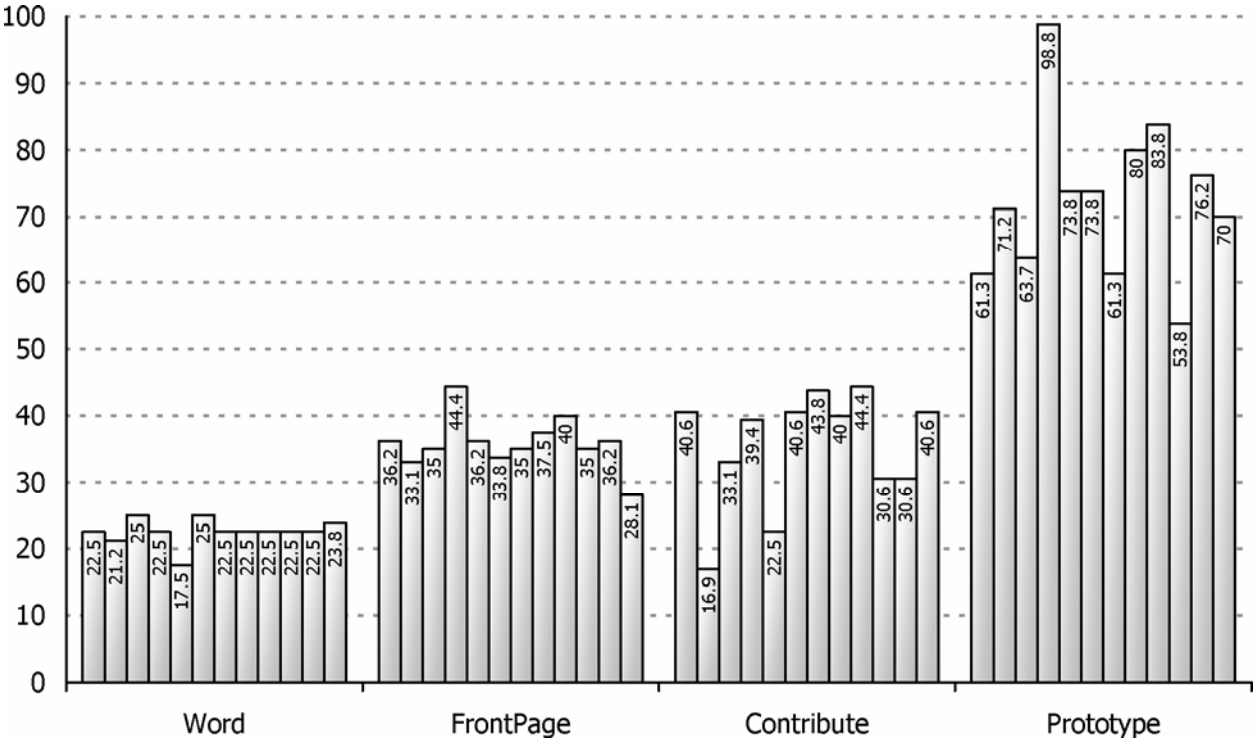
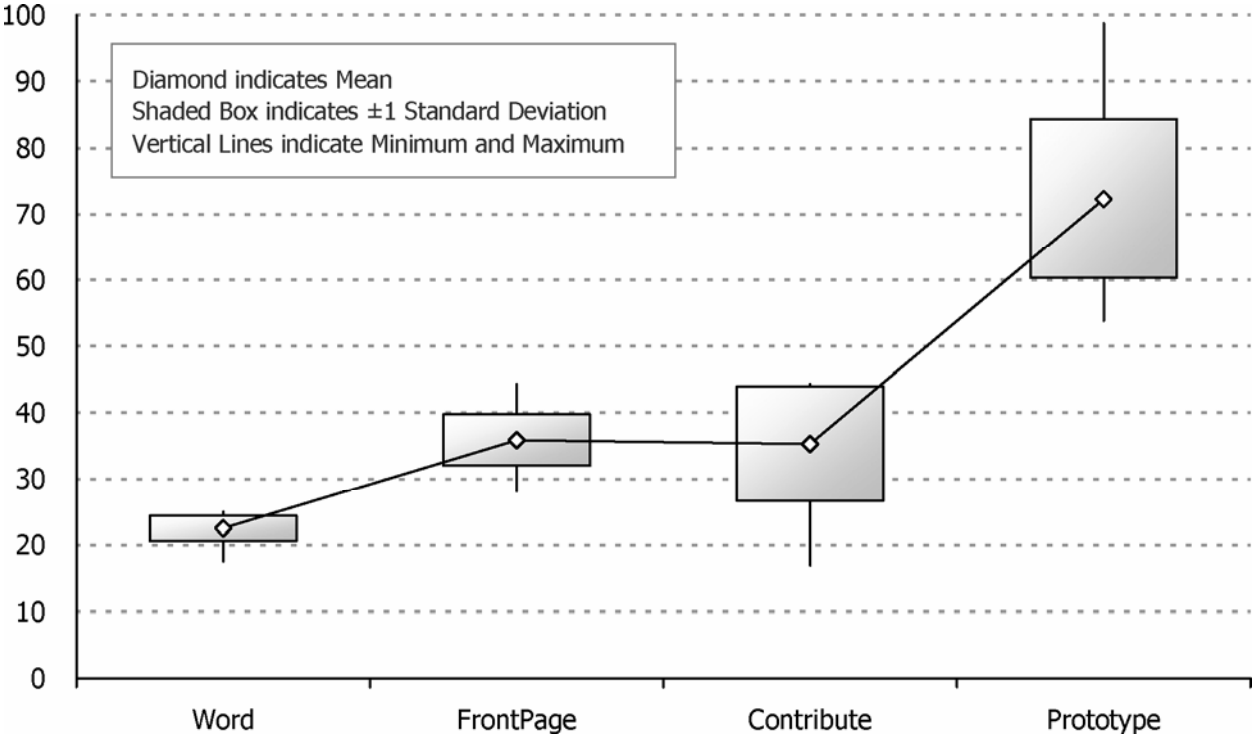


Figure 24. User Test Accessibility Scores Box Plot (%)



Element Scores

Figure 25 shows the average accessibility scores for each tool broken down by element. The results reveal that, even with the prototype, there is significant room for improvement in dealing with specific elements.

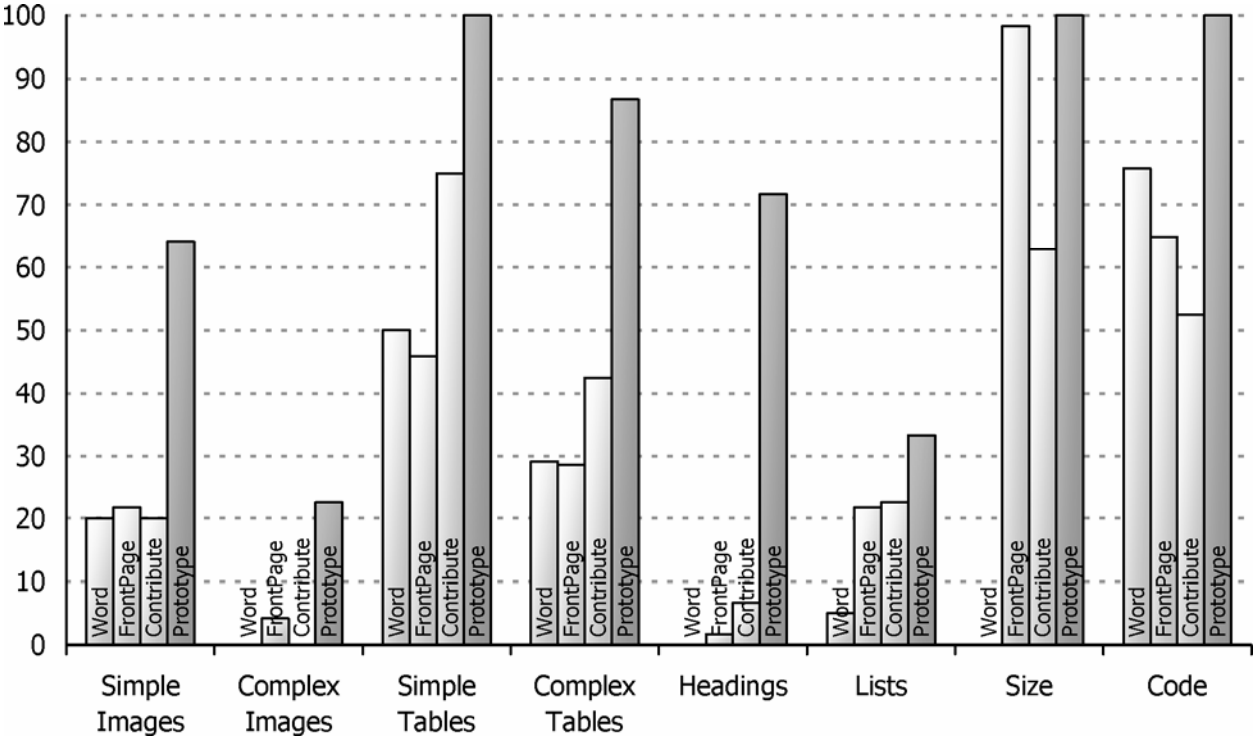
Complex images, those requiring alternate text and a full description, were the most problematic, with the prototype achieving just 22.5% and the other tools effectively failing completely. In observations, it was clear that even with prompting from the prototype, users did not understand why or how to add a full description for the complex image.

Lists were also surprisingly difficult, with the prototype averaging only 33.3%. Observations revealed that most users attempted to create lists by typing bullets or numbers and using spaces to indent sub-lists; most users did not use list features in any tool, even when they were displayed prominently on the menu or toolbar.

Simple images, those requiring only alternate text, were problematic for a different reason – several users identified the simple image correctly and then re-classified it as decorative after being prompted to enter alternate text, apparently in an active effort to avoid having to enter alternate text.

The prototype performed expectedly well with simple tables (100%), which it fully automated, and headings (71.7%), for which the Heading Check provided assistance.

Figure 25. User Test Accessibility Scores by Element (%)



Ease-of-Use Ratings

On the post-test questionnaire, users were asked to respond to the following question for each of the tools tested:

> Please rate [tool name] on ease of use, using a scale from 1 (very difficult) to 5 (very easy).

Figure 26 and Figure 27 show the responses to this question. Interestingly, there is only minor variation between the tools, with Word averaging 3.7, the prototype 3.4, and Contribute and FrontPage 3.3 and 2.9 respectively. ANOVA calculations (Table A-8 in the Appendix) show that there is no significant difference between the tools at the 95% confidence level. This observation is particularly interesting when considering that users reported having an average of 9.3 years of experience with Word.

Figure 26. User Ease-of-Use Ratings

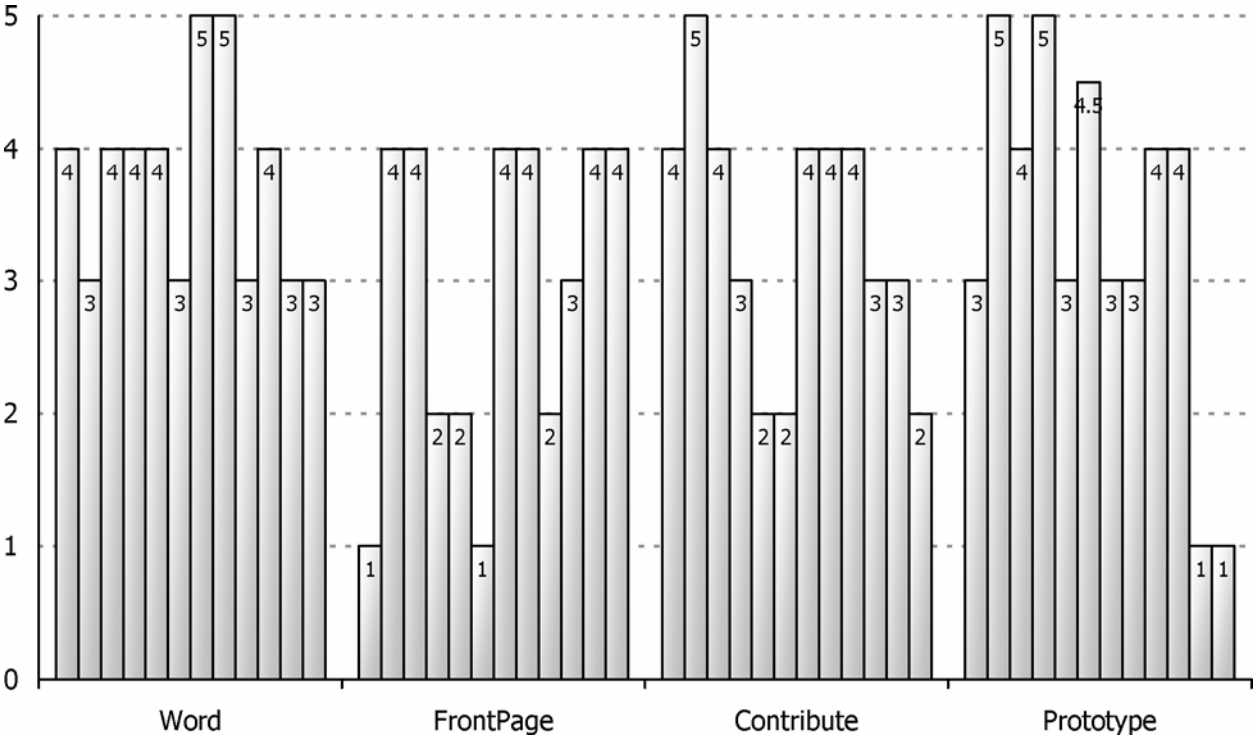


Figure 27. Average User Ease-of-Use Ratings (1-5)

Tool	Average Rating	Minimum Rating	Maximum Rating	Standard Deviation	Number of Users
Word	3.7	3	5	0.75	12
FrontPage	2.9	1	4	1.24	12
Contribute	3.3	2	5	0.98	12
Prototype	3.4	1	5	1.33	12

Observations

While observing and scoring the user tests, several interesting trends in user behavior became evident. These trends will inform future improvements made to the "accessibility first" authoring tool prototype.

- While all four tools tested provided means to enter alternate text for images, only one user entered alternate text in any tool other than the prototype (and he did so only after using the prototype).
- In the prototype, four users set the image meaning to "decorative" for the complex image in order to avoid having to enter alternate text.
- Although the prototype prompted users to enter a full description for the complex image, only one user did so.
- Only half of the users selected row/column headers in Contribute, even though the option was clearly available on the insert table dialog.
- Only one user used headings in any tool other than the prototype. Even in the prototype, four users used bold instead of headings.
- In Word, six users attempted to create bulleted lists by inserting bullet symbols.

These observations clearly reinforce the belief that accessibility needs to be automated wherever possible, and where it cannot be automated, users needs to be actively prompted and provided extremely clear and direct instruction.

User Feedback

Several of the user test participants supplied encouraging feedback regarding the "accessibility first" authoring tool prototype in their post-test questionnaires:

- I liked the ability to identify things for a person with disabilities.
- Though I personally don't like going through the trouble of accessibility for everyone, I think it's good.
- With its interface asking simple questions, it would be easy to use even if you don't know the lingo.
- It really was incredibly easy to use.

Expert Tests

Results of the expert tests were very much in line with those of the user tests, again confirming the two hypotheses proposed in this study.

Figure 28, which compares the expert "typical use" test scores to the user test average scores, shows a strong agreement between the two, suggesting that the expert typical use guidelines provided an effective means of simulating typical use.

Figure 28. Expert "Typical Use" & User Test Accessibility Scores (%)

Tool	Expert "Typical Use" Score	User Test Average Score	User Test Minimum Score	User Test Maximum Score	User Test Standard Deviation
Word	25	22.5	17.5	25	1.9
FrontPage	36.2	35.9	28.1	44.4	3.9
Contribute	35.6	35.3	16.9	44.4	8.7
Word, FrontPage, Contribute	32.3	31.2	16.9	44.4	8.3
Prototype	77.5	72.3	53.8	98.8	12.0

Expert test scores for the remaining tools and the "best case" scenario are shown in Figure 29 and Figure 30. Of the additional tools tested, few performed substantially better than those already tested, and none approached the performance of the prototype: Together, the non-technical tools (Word, WordPerfect, OpenOffice, Contribute, eWebEditPro, r.a.d.editor) averaged 34.7%, and the technical tools (FrontPage, Dreamweaver, XStandard) averaged 46.9%. The prototype scored 77.5%.

Notably, even the technical authoring tools that included accessibility features – FrontPage, Dreamweaver and XStandard – still failed to produce adequately accessible documents in the "typical use" scenario. "Best case" testing of these tools, however, revealed that accessibility was certainly possible, averaging 95%. This observation led to an important additional conclusion:

> It is not adequate to make accessibility ***possible***, it must be ***automatic***.

Figure 29. Expert Test Accessibility Scores (%)

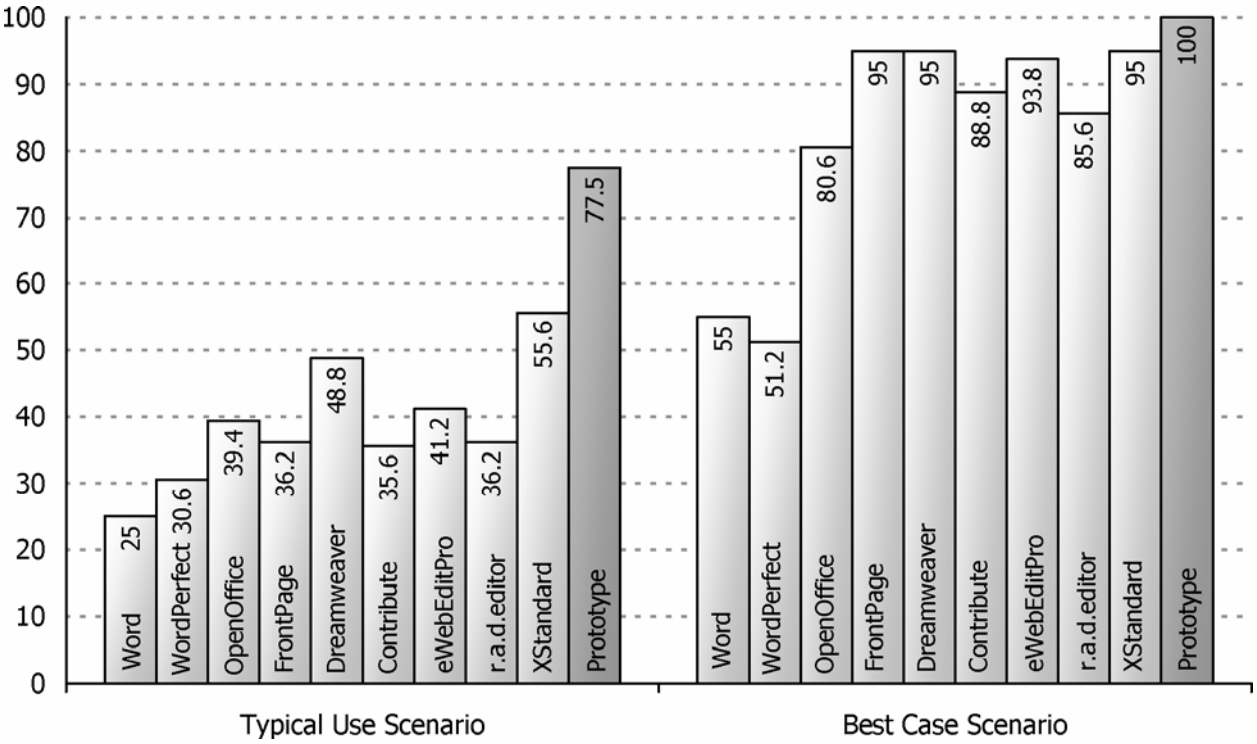


Figure 30. Expert Test Accessibility Scores (%)

Tool	Typical Use	Best Case
Word	25	55
WordPerfect	30.6	51.2
OpenOffice	39.4	80.6
FrontPage	36.2	95
Contribute	35.6	88.8
Dreamweaver	48.8	95
eWebEditPro	41.2	93.8
r.a.d.editor	36.2	85.6
XStandard	55.6	95
Non-Technical Tools (Word, WordPerfect, OpenOffice, Contribute, eWebEditPro, r.a.d.editor)	34.7	75.8
Technical Tools (FrontPage, Dreamweaver, XStandard)	46.9	95
Prototype	77.5	100

Conclusions

The results of Phase I user and expert tests confirmed the project's two main hypotheses:

(1) Existing authoring tools do not adequately enable typical authors to create accessible documents.

(2) It is possible for an authoring tool designed to promote accessibility to significantly improve the accessibility of documents created by typical authors.

In addition, the results of Phase I also lead the project team to two other important conclusions:

(3) There is still significant room to improve the effectiveness of the "accessibility first" authoring tool prototype.

(4) It is not adequate to make accessibility possible; it must be automatic.

Based on the results of Phase I, the project team is convinced that it is both necessary and possible to continue the development of the "accessibility first" prototype with the ultimate goal of producing an authoring tool that will enable non-technical users to easily and reliably create highly accessible documents.

Appendix

Table A-1. User Test Accessibility Score Data, Word

	Simple Images		Complex Images		Simple Tables	Complex Tables	Headings					Lists					Size	Code		Total	Percent
User	1	2	Alt	Desc	1	2	1	2	3	4	5	1	2	3	4	5	Size	HTML	CSS		
1	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
2	0	2	0	0	5	2	0	0	0	0	0	0	0	0	0	0	0	4	4	17	21.3%
3	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	0	4	4	20	25.0%
4	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
5	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0	4	14	17.5%
6	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	0	4	4	20	25.0%
7	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
8	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
9	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
10	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
11	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	4	4	18	22.5%
12	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	0	4	3	19	23.8%
Sub-Total	0	24	0	0	60	35	0	0	0	0	0	6	0	0	0	0	0	44	47	216	
Total	24		0		60	35	0					6					0	91		216	
Average	2.00		0.00		5.00	2.92	0.00					0.50					0.00	7.58		18.00	
Percent	20.0%		0.0%		50.0%	29.2%	0.0%					5.0%					0.0%	75.8%		22.5%	

Table A-2. User Test Accessibility Score Data, FrontPage

	Simple Images		Complex Images		Simple Tables	Complex Tables	Headings					Lists					Size	Code		Total	Percent
User	1	2	Alt	Desc	1	2	1	2	3	4	5	1	2	3	4	5	Size	HTML	CSS		
1	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	3	29	36.3%
2	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	9.5	4	3	26.5	33.1%
3	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	2	28	35.0%
4	0	2	0	0	5	3	0	0	0	0	0	2	2	2	2	2	8.5	4	3	35.5	44.4%
5	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	3	29	36.3%
6	0	2	0	0	5	3	0	0	0	0	0	0	0	0	0	0	10	4	3	27	33.8%
7	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	2	28	35.0%
8	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	4	30	37.5%
9	2	2	5	0	5	3	2	0	0	0	0	0	0	0	0	0	10	2	1	32	40.0%
10	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	2	28	35.0%
11	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	3	29	36.3%
12	0	2	0	0	0	1.5	0	0	0	0	0	2	0	0	0	0	10	4	3	22.5	28.1%
Sub-Total	2	24	5	0	55	34.5	2	0	0	0	0	18	2	2	2	2	118	46	32	344.5	
Total	26		5		55	34.5	2					26					118	78		344.5	
Average	2.17		0.42		4.58	2.88	0.17					2.17					9.83	6.50		28.71	
Percent	21.7%		4.2%		45.8%	28.7%	1.7%					21.7%					98.3%	65.0%		35.9%	

Table A-3. User Test Accessibility Score Data, Contribute

	Simple Images		Complex Images		Simple Tables	Complex Tables	Headings					Lists					Size	Code		Total	Percent
User	1	2	Alt	Desc	1	2	1	2	3	4	5	1	2	3	4	5	Size	HTML	CSS		
1	0	2	0	0	10	6	0	0	0	0	0	2	0	0	0	0	7.5	0	5	32.5	40.6%
2	0	2	0	0	5	1.5	0	0	0	0	0	0	0	0	0	0	0	0	5	13.5	16.9%
3	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	9.5	0	5	26.5	33.1%
4	0	2	0	0	5	3	0	0	0	0	0	2	2	1	2	2	7.5	0	5	31.5	39.4%
5	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	1	0	5	18	22.5%
6	0	2	0	0	10	6	0	0	0	0	0	2	0	0	0	0	7.5	0	5	32.5	40.6%
7	0	2	0	0	10	4.5	0	0	0	0	0	2	0	0	0	0	8.5	4	4	35	43.8%
8	0	2	0	0	10	6	0	0	0	0	0	0	0	0	0	0	10	0	4	32	40.0%
9	0	2	0	0	10	6	0	2	2	2	2	2	0	0	0	0	0.5	4	3	35.5	44.4%
10	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	8.5	0	4	24.5	30.6%
11	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	7.5	0	5	24.5	30.6%
12	0	2	0	0	10	6	0	0	0	0	0	2	0	0	0	0	7.5	0	5	32.5	40.6%
Sub-Total	0	24	0	0	90	51	0	2	2	2	2	20	2	1	2	2	75.5	8	55	338.5	
Total	24		0		90	51	8					27					75.5	63		338.5	
Average	2.00		0.00		7.50	4.25	0.67					2.25					6.29	5.25		28.21	
Percent	20.0%		0.0%		75.0%	42.5%	6.7%					22.5%					62.9%	52.5%		35.3%	

Table A-4. User Test Accessibility Score Data, Prototype

	Simple Images		Complex Images		Simple Tables	Complex Tables	Headings					Lists					Size	Code		Total	Percent
User	1	2	Alt	Desc	1	2	1	2	3	4	5	1	2	3	4	5	Size	HTML	CSS		
1	0	5	0	0	10	10	2	0	0	0	0	2	0	0	0	0	10	5	5	49	61.3%
2	0	5	0	0	10	10	2	2	2	2	2	2	0	0	0	0	10	5	5	57	71.3%
3	2	4	5	0	10	6	2	0	0	0	0	2	0	0	0	0	10	5	5	51	63.8%
4	5	5	5	4	10	10	2	2	2	2	2	2	2	2	2	2	10	5	5	79	98.8%
5	4	5	2	0	10	6	2	2	2	2	2	2	0	0	0	0	10	5	5	59	73.8%
6	2	4	2	0	10	10	2	1	2	2	2	2	0	0	0	0	10	5	5	59	73.8%
7	0	5	0	0	10	10	2	0	0	0	0	2	0	0	0	0	10	5	5	49	61.3%
8	2	5	5	0	10	10	2	2	2	2	2	2	0	0	0	0	10	5	5	64	80.0%
9	2	5	0	0	10	10	2	2	2	2	2	2	2	2	2	2	10	5	5	67	83.8%
10	0	2	0	0	10	6	2	0	0	0	1	2	0	0	0	0	10	5	5	43	53.8%
11	5	4	2	0	10	10	2	2	2	0	2	2	0	0	0	0	10	5	5	61	76.3%
12	2	4	2	0	10	6	2	2	2	2	2	2	0	0	0	0	10	5	5	56	70.0%
Sub-Total	24	53	23	4	120	104	24	15	16	14	17	24	4	4	4	4	120	60	60	694	
Total	77		27		120	104	86					40					120	120		694	
Average	6.42		2.25		10.00	8.67	7.17					3.33					10.00	10.00		57.83	
Percent	64.2%		22.5%		100.0%	86.7%	71.7%					33.3%					100.0%	100.0%		72.3%	

Table A-5. User Test Accessibility Scores ANOVA, Single Factor

User Test Accessibility Scores (%)

User	Word	FrontPage	Contribute	Prototype
1	22.5	36.2	40.6	61.3
2	21.2	33.1	16.9	71.2
3	25	35	33.1	63.7
4	22.5	44.4	39.4	98.8
5	17.5	36.2	22.5	73.8
6	25	33.8	40.6	73.8
7	22.5	35	43.8	61.3
8	22.5	37.5	40	80
9	22.5	40	44.4	83.8
10	22.5	35	30.6	53.8
11	22.5	36.2	30.6	76.2
12	23.8	28.1	40.6	70

Summary

Tools	Count	Sum	Average	Variance
Word	12	270.0	22.5	3.7
FrontPage	12	430.5	35.9	15.2
Contribute	12	423.1	35.3	75.4
Prototype	12	867.7	72.3	143.5

ANOVA: Single Factor

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P	F critical
Between Tools	16569.0	3	5523.0	92.890	0.000	2.816
Within Tools	2616.1	44	59.5			
Total	19185.1	47				

ANOVA calculations performed using Microsoft Excel 2003 Analysis ToolPak

Table A-6. User Test Accessibility Scores ANOVA, Two-Factor

User Test Accessibility Scores by Test Order (%)

Test Order	Word	FrontPage	Contribute	Prototype
1	22.5	44.4	43.8	53.8
1	21.2	36.2	40	76.2
1	25	33.8	44.4	70
2	22.5	36.2	16.9	63.7
2	22.5	37.5	22.5	73.8
2	22.5	36.2	40.6	83.8
3	17.5	35	40.6	71.2
3	22.5	35	40.6	98.8
3	22.5	28.1	30.6	80
4	25	33.1	33.1	61.3
4	22.5	40	39.4	73.8
4	23.8	35	30.6	61.3

ANOVA: Two-Factor with Replication

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P	F critical
Test Order	125.8	3	41.9	0.869	0.467	2.901
Tool	16569.0	3	5523.0	114.419	0.000	2.901
Interaction	945.7	9	105.1	2.177	0.051	2.189
Within	1544.6	32	48.3			
Total	19185.1	47				

ANOVA calculations performed using Microsoft Excel 2003 Analysis ToolPak

Table A-7. User Survey Data

User	Tool	Experience (years)	Frequency (hours/week)	Proficiency (1-5)	Satisfaction (1-5)
1	Word	7.5	2	3	4
1	FrontPage	0.5	0	1	1
1	Contribute	0	0	0	4
1	Prototype	0	0	0	3
2	Word	8	2	3	3
2	FrontPage	0	0	0	4
2	Contribute	0	0	0	5
2	Prototype	0	0	0	5
3	Word	2	0	2	4
3	FrontPage	0	0	0	4
3	Contribute	0	0	0	4
3	Prototype	0	0	0	4
4	Word	6.5	1	4	4
4	FrontPage	0	0	0	2
4	Contribute	0	0	0	3
4	Prototype	0	0	0	5
5	Word	14	10	3	4
5	FrontPage	0.5	0	1.5	2
5	Contribute	0	0	0	2
5	Prototype	0	0	0	3
6	Word	10	22.5	3.5	3
6	FrontPage	0	0	1.5	1
6	Contribute	0	0	0	2
6	Prototype	0	0	0	4.5

Continued on next page

Table A-7. User Survey Data, continued

User	Tool	Experience (years)	Frequency (hours/week)	Proficiency (1-5)	Satisfaction (1-5)
7	Word	14	7.5	4	5
7	FrontPage	0	0	0	4
7	Contribute	0	0	0	4
7	Prototype	0	0	0	3
8	Word	10	30	3.5	5
8	FrontPage	0	0	0	4
8	Contribute	0	0	0	4
8	Prototype	0	0	0	3
9	Word	4.5	2.5	3	3
9	FrontPage	3	0	3	2
9	Contribute	0	0	0	4
9	Prototype	0	0	0	4
10	Word	10	2	2	4
10	FrontPage	0	0	0	3
10	Contribute	0	0	0	3
10	Prototype	0	0	0	4
11	Word	10	30	4	3
11	FrontPage	0	0	0	4
11	Contribute	0	0	0	3
11	Prototype	0	0	0	1
12	Word	15	25	3	3
12	FrontPage	0	0	0	4
12	Contribute	0	0	0	2
12	Prototype	0	0	0	1

Table A-8. User Survey Satisfaction ANOVA, Single Factor

User Survey Satisfaction Scores (1-5)

User	Word	FrontPage	Contribute	Prototype
1	4	1	4	3
2	3	4	5	5
3	4	4	4	4
4	4	2	3	5
5	4	2	2	3
6	3	1	2	4.5
7	5	4	4	3
8	5	4	4	3
9	3	2	4	4
10	4	3	3	4
11	3	4	3	1
12	3	4	2	1

Summary

Tool	Count	Sum	Average	Variance
Word	12	45.0	3.8	0.6
FrontPage	12	35.0	2.9	1.5
Contribute	12	40.0	3.3	1.0
Prototype	12	40.5	3.4	1.8

ANOVA: Single Factor

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	P	F critical
Between Tools	4.2	3	1.4	1.149	0.340	2.816
Within Tools	53.4	44	1.2			
Total	57.6	47				

ANOVA calculations performed using Microsoft Excel 2003 Analysis ToolPak

Table A-9. Expert Test Accessibility Score Data

	Simple Images		Complex Images		Simple Tables	Complex Tables	Headings					Lists					Size	Code		Total	Percent
Tool	1	2	Alt	Desc	1	2	1	2	3	4	5	1	2	3	4	5	Size	HTML	CSS		
Typical Use Scenario																					
Word	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	0	4	4	20	25.0%
WordPerfect	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	7.5	0	5	24.5	30.6%
OpenOffice	0	2	0	0	7.5	4.5	0	0	0	0	0	2	0	0	0	0	8.5	3	4	31.5	39.4%
FrontPage	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	4	3	29	36.2%
Contribute	0	2	0	0	5	3	0	0	0	0	0	2	0	0	0	0	7.5	4	5	28.5	35.6%
Dreamweaver	0	2	0	0	10	6	0	0	0	0	0	2	0	0	0	0	10	4	5	39	48.8%
eWebEditPro	0	4	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	5	4	33	41.2%
r.a.d.editor	0	5	0	0	5	3	0	0	0	0	0	2	0	0	0	0	10	0	4	29	36.2%
XStandard	2	5	2	0	7.5	6	0	0	0	0	0	2	0	0	0	0	10	5	5	44.5	55.6%
Prototype	2	5	2	1	10	10	2	2	2	2	2	2	0	0	0	0	10	5	5	62	77.5%
Best Case Scenario																					
Word	5	2	5	4	5	3	2	2	2	2	2	2	0	0	0	0	0	4	4	44	55.0%
WordPerfect	0	2	0	4	5	3	2	2	2	2	2	2	0	0	0	0	10	0	5	41	51.2%
OpenOffice	5	2	5	4	7.5	4.5	2	2	2	2	2	2	2	2	2	2	8.5	3	5	64.5	80.6%
FrontPage	5	5	5	5	10	6	2	2	2	2	2	2	2	2	2	2	10	5	5	76	95.0%
Contribute	5	2	5	4	10	6	2	2	2	2	2	2	2	2	2	2	10	4	5	71	88.8%
Dreamweaver	5	5	5	5	10	6	2	2	2	2	2	2	2	2	2	2	10	5	5	76	95.0%
eWebEditPro	5	5	5	4	10	6	2	2	2	2	2	2	2	2	2	2	10	5	5	75	93.8%
r.a.d.editor	5	5	5	5	10	4.5	2	2	2	2	2	2	2	2	2	2	10	0	4	68.5	85.6%
XStandard	5	5	5	5	10	6	2	2	2	2	2	2	2	2	2	2	10	5	5	76	95.0%
Prototype	5	5	5	5	10	10	2	2	2	2	2	2	2	2	2	2	10	5	5	80	100.0%